Student: Yixin Chang (1005991651)
INF2178 - Assignment 4

Analyzing MMSE Score Trends in Dementia

1. Introduction

Dementia, which affects millions of people worldwide, severely disrupts memory and cognitive processes, emphasizing the significance of continued study for better understanding and treatment. The dataset "INF2178_A4_data.csv" collects data from a study on MRI results of patients with/without dementia. This report focuses on changes in certain indices over time in individuals with and without dementia.
Our exploration will address one research question:

- **Research Question:** How do Mini Mental State Examination (MMSE) scores change over time across dementia statuses (non-demented, demented, converted)?

2. Data Cleaning and Data Wrangling

We reduce our dataset to the following columns:

- Subject ID: Identifier for the subject.
- Group: Nondemented, Demented, or Converted.
- Visit: Visit Order.
- MMSE: Mini Mental State Examination.

The current dataset has 294 rows and 4 columns. After critically reviewing the data, we found that there is one missing value (NaN) in the MMSE column. We use the mean value of the MMSE column to fill this missing value, then there are no missing values in the dataset. We will work on our current dataset for further analysis.

3. Exploratory Data Analysis (EDA)

We perform exploratory data analysis to visualize data distribution. The following table presents a descriptive statistical summary of the Mini-Mental State Examination (MMSE) scores from the dataset.

count	mean	std	min	25%	50%	75%	max
294	27.259	3.408	15	26	29	30	30

Figure 1: Summary of MMSE

The range of MMSE is 15-30, and the mean is 27.259. From this analysis, we can infer that the majority of individuals in this dataset have MMSE scores that suggest minimal to no cognitive impairment.

Then, we will see if the distribution of MMSE scores has difference among various groups (Nondemented, Demented, or Converted), and Figure 2 shows these distributions using histograms and boxplots. From the left histogram, the Nondemented group peaks around the maximum score of 30, which indicates good cognitive function. The Demented group has a wider distribution, with most scores lower than the Nondemented group. From the right boxplot, the median MMSE score of the Nondemented group is near the maximum, and its IQR is small, indicating most scores are high and close together. The Demented group has a lower median and a larger IQR, suggesting greater variability in scores and more individuals with lower cognitive function.

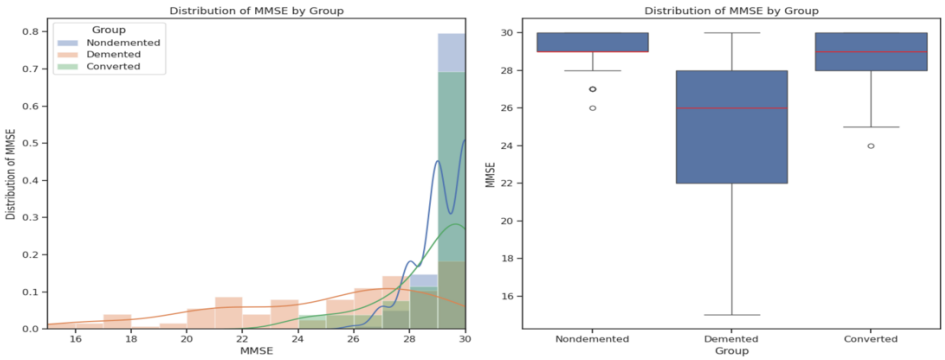
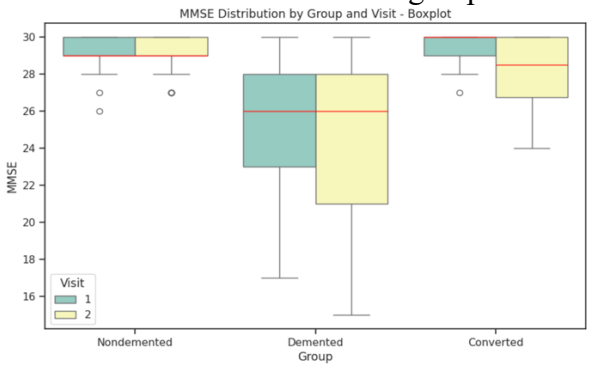


Figure 2: Histogram (left) and Boxplots (right) for MMSE by Group

In the next step, boxplots and a descriptive analysis are used in Figure 3 to compare the distribution of MMSE scores across groups over time by considering visit order.



		mean	std
Visit	Group		
1	Converted	29.36	0.93
	Demented	25.33	3.32
	Nondemented	29.19	0.85
2	Converted	28.00	2.09
	Demented	24.29	4.38
	Nondemented	29.11	0.96

Figure 3: Boxplots (left) and Summary (right) for MMSE scores by Group and Visit

The boxplot and statistical summary show that the Nondemented group's MMSE scores remained stable after two visits, whereas the Converted group shows a notable cognitive decline. Moreover, the Demented group demonstrates slight worsening, indicating increasing cognitive impairment.

We also use Figure 4 to see the trend of MMSE scores with different groups and visit order.

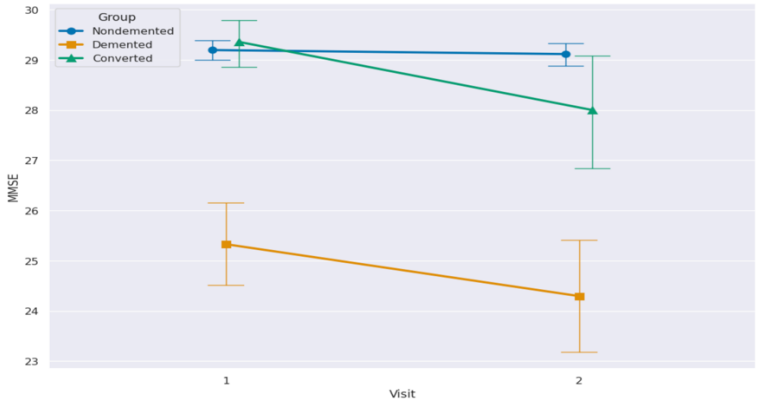


Figure 4: Point plot shows the mean MMSE scores for three groups (Nondemented, Demented, Converted) over two visits (1/2)

The line of the Nondemented group is almost horizontal with a high score, indicating stable cognitive function between two visits. Both the Demented and Converted group's scores decrease significantly from the first to the second visit, so the cognitive function is declined.

Furthermore, the spread of scores becomes larger for these two groups, and this may imply the increase of variability.

4. Mixed-effects ANOVA

Research Question: How do Mini Mental State Examination (MMSE) scores change over time across dementia statuses (non-demented, demented, converted)?

Considering our research question, we will use mixed-effects ANOVA to explore if there are significant differences of MMSE score distribution among various group from Visit 1 to Visit 2. In this case, the dependent variable is MMSE score, the within-subject factor is visit order (Visit), and the between-subject factor consists of groups (Nondemented, Demented, Converted). Figure 5 shows the results of ANOVA summary.

Source	SS	DF1	DF2	MS	F	p-unc	np2	eps
Group	1322.017	2	141	661.009	56.100	< 0.001	0.443	nan
Visit	21.528	1	141	21.528	8.525	0.004	0.057	1
Interaction	16.204	2	141	8.102	3.208	0.043	0.044	nan

Figure 5: Mixed-effects ANOVA summary

From the results, all three p-values are less than 0.05 (significance level), so we reject the null hypothesis. In other words, there is a strong difference in MMSE scores among the groups, and the time factor of visits influences MMSE scores. The interaction between Group and Visit is also significant ($p = 0.043$) with an effect size of 0.044, suggesting the impact of visit order on MMSE scores differs between groups.

We also perform post hoc tests in Figure 6 to see more details.

Contrast	Visit	A	B	Paired	Parametric	T	dof	alternative	p-unc	BF10	hedges
Visit	-	1	2	True	True	2.876	143.000	two-sided	0.005	4.793	0.159
Group	-	Converted	Demented	False	True	6.745	50.480	two-sided	< 0.001	$2.189 * 10^{6}$	1.164
Group	-	Converted	Nondemented	False	True	-1.303	12.315	two-sided	0.216	0.599	-0.584
Group	-	Demented	Nondemented	False	True	-9.512	65.514	two-sided	< 0.001	$3.306 * 10^{13}$	-1.745
Visit * Group	1	Converted	Demented	False	True	8.076	60.165	two-sided	< 0.001	$4.479 * 10^{8}$	1.319
Visit * Group	1	Converted	Nondemented	False	True	0.489	13.999	two-sided	0.633	0.336	0.167
Visit * Group	1	Demented	Nondemented	False	True	-9.124	68.185	two-sided	< 0.001	$3.993 * 10^{12}$	-1.669
Visit * Group	2	Converted	Demented	False	True	4.515	33.372	two-sided	< 0.001	693.697	0.891
Visit * Group	2	Converted	Nondemented	False	True	-1.816	11.802	two-sided	0.095	1.126	-0.937
Visit * Group	2	Demented	Nondemented	False	True	-8.480	66.145	two-sided	< 0.001	$1.265 * 10^{11}$	-1.554

Figure 6: Post hoc tests results

The post hoc tests show significant differences between the Converted and Demented groups ($p < 0.001$), as well as the Demented and Nondemented groups ($p < 0.001$), but not between Converted and Nondemented groups ($p = 0.216 > 0.05$). Interactions between the Converted and Demented groups at both visits ($p < 0.001$) are significant, as well as interactions between the Demented and NonDemented groups. This means that the rate of change in MMSE scores over time differs between these groups. However, interactions between the Converted and Nondemented groups at both visits are not significant with $p = 0.633 > 0.05$ and $p = 0.095 > 0.05$ respectively.

Checking assumptions is an important step when performing mixed-effects ANOVA. To do this, we firstly use Mauchly's test of sphericity to check whether the variances of the differences among all combinations of related groups are equal. The result of the test 1.0, which indicates perfect sphericity. This means that the assumption of sphericity has been met. Secondly, a normality test is performed, and the test statistic used is Shapiro-Wilk test. The results are displayed in Figure 7. For both visits, we find that p-value < 0.001 and the result is False. We can conclude that the MMSE score distributions for Visit 1 and Visit 2 are not

from a normal distribution, which violates the normality assumption. This may influence the accuracy of the results.

Visit	W	pval	normal
1	0.787	< 0.001	False
2	0.761	< 0.001	False

Figure 7: Results for the normality test

Lastly, because the data is not drawn from normal distribution, Levene's test is used to check the assumption, the Homogeneity of variances. Figure 8 shows the results for this test.

	W	pval	Equal_var
levene	64.643	< 0.001	False

Figure 8: Results for Levene' test

As the p value (< 0.05) is significant in Figure 8, we reject null hypothesis and conclude the assumption of homogeneity of variances has been violated.

5. Power Analysis Plot

In this part, we create a statistical power analysis plot for t-tests with given effect size = 0.7 in Figure 9.

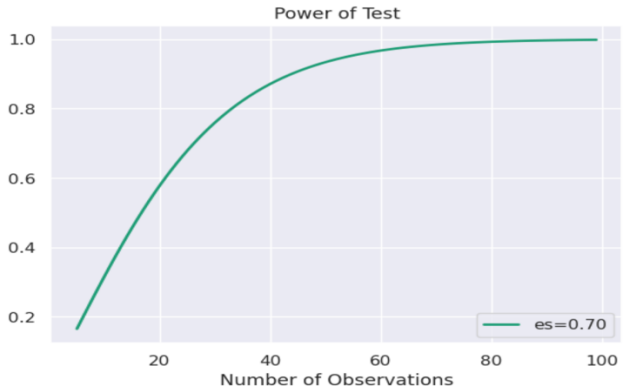


Figure 9: Power Plot with effect size = 0.7

Statistical power is the probability of successfully rejecting the null hypothesis, and this plot visualizes how it changes with different sample sizes, given effect size = 0.7. As shown, with an increasing number of observations, the power increases, approaching to 1 as the number of observations approaches 100. We also use TTestIndPower to find required sample size in each group for a two-sided t-test. Given that power = 0.91, alpha = 0.05, and effect size = 0.7, the result sample size is 45.451. Since the number of individuals should be integer, we would need to include at least 46 participants.

6. Conclusion

In conclusion, MMSE scores differ significantly across dementia statuses and change over time. Using MMSE as the measurement approach, a decreasing trend in cognitive functions is observed for individuals with demented status. In addition, for people who transition from nondemented to demented status, the decline is even more pronounced as time goes by. However, the violation of the normality assumption and homogeneity of variances may cause incorrect results, suggesting that other methods may be used to validate the results.